An iron-sulfur cluster in the family 4 uracil-DNA glycosylases

Hinks, J A ; Evans, M C W ; De Miguel, Y ; Sartori, Alessandro A ; Jiricny, J ; Pearl, L H

Abstract: The 25-kDa Family 4 uracil-DNA glycosylase (UDG) from Pyrobaculum aerophilum has been expressed and purified in large quantities for structural analysis. In the process we observed it to be colored and subsequently found that it contained iron. Here we demonstrate that P. aerophilum UDG has an iron-sulfur center with the EPR characteristics typical of a 4Fe4S high potential iron protein. Interestingly, it does not share any sequence similarity with the classic iron-sulfur proteins, although four cysteines (which are strongly conserved in the thermophilic members of Family 4 UDGs) may represent the metal coordinating residues. The conservation of these residues in other members of the family suggest that 4Fe4S clusters are a common feature. Although 4Fe4S clusters have been observed previously in Nth/MutY DNA repair enzymes, this is the first observation of such a feature in the UDG structural superfamily. Similar to the Nth/MutY enzymes, the Family 4 UDG centers probably play a structural rather than a catalytic role.

DOI: https://doi.org/10.1074/jbc.M200668200

DNA Replication Repair and Recombination:

An iron-sulphur cluster in the Family-4uracil-DNA glycosylases

John A. Hinks, Michael C. W. Evans, Yolanda de Miguel, Alessandro A. Sartori, Josef Jiricny and Laurence H. Pearl

J. Biol. Chem. published online February 27, 2002

JBC Papers in Press. Published on February 27, 2002 as Manuscript M200668200

An iron-sulphur cluster in the Family-4 uracil-DNA glycosylases.

John A. Hinks[1], Michael C.W. Evans[2], Yolanda de Miguel[3], Alessandro A. Sartori[4], Josef Jiricny[4], and Laurence H. Pearl[1§]

1. CRC DNA Repair Enzyme Group, Section of Structural Biology, The Institute of Cancer Research, 237 Fulham Road, London , SW3 6JB, UK
2. Department of Biology, Darwin Building, University College London, Gower St., London WC1E 6BT, U.K.
3. Department of Chemistry, Kings College London, The Strand, London, WC2R 2LS, U.K.
4. Institute of Medical Radiobiology, University of Zurich, Switzerland.

§ Corresponding author.

Telephone: 020 7970 6046 Fax: 020 7970 6051 email: l.pearl@icr.ac.uk

Running title: Iron sulphur centre in uracil-DNA glycosylases.

SUMMARY

The 25 kD Family-4 uracil-DNA glycosylase (UDG) from *Pyrobaculum aerophilum* has been expressed and purified in large quantities for structural analysis. In the process we observed it to be coloured and subsequently found that it contained iron. Here we demonstrate that it has an iron-sulphur centre with the EPR characteristics typical of a 4Fe4S High Potential Iron Protein. Interestingly, it does not share any sequence similarity with the classic iron-sulphur proteins, though four cysteines, which are strongly conserved in the thermophilic members of Family 4 UDGs, may represent the metal co-ordinating residues. The conservation of these residues in other members of the family suggest that 4Fe4S clusters are a common feature. Although 4Fe4S clusters have previously been observed in MutY/Nth DNA repair enzymes, this is the first observation of such a feature in the UDG structural super-family. As in the Nth/MutY enzymes, the Family-4 UDG centres probably play a structural, rather than catalytic, role.

INTRODUCTION

Uracil-DNA glycosylases are ubiquitous DNA repair enzymes responsible for the excision of uracil bases from DNA as the first step in a base excision repair pathway. Uracil arises in DNA either as a result of the hydrolytic deamination of cytosine residues in G:C base-pairs (1), or due to incorporation of deoxyuridine monophosphate (instead of thymidine monophosphate) opposite adenine during DNA replication (2). If left uncorrected, the former process would cause G:C to A:T transition mutations (1), while the latter could result in the disruption of specific regulatory DNA-protein interactions (3).

Hyperthermophilic organisms are at especially high risk of DNA damage by cytosine deamination, which is significantly enhanced by elevated temperature (4). Since hyperthermophiles do not exhibit any greater susceptibility to this type of damage they presumably possess more effective repair enzymes (5). However, despite the detection of UDG activity in several hyperthermophiles (6) no sequences homologous to the archetypal *E.coli ung*-encoded enzyme were initially apparent in archaeal genomes. Subsequently, UDGs were identified in hyperthermophilic eubacteria and archaea (7-9) with more obvious homology to a second family of uracil base-excision repair enzymes typified by the human thymine DNA glycosylase (TDG) (10) and the bacterial MUG (11). These G:T/U mismatch-specific enzymes (Family-2) are structurally and mechanistically related to the UNG-type UDGs (Family-1) (12,13) and unite the UNG-type and thermophile enzymes (Family-4) into a uracil-DNA glycosylase superfamily (14).

Pyrobaculum aerophilum is a hyperthermophilic archaeon isolated from a boiling marine water hole, and growing optimally at 100°C and pH 7.0 (15). A fosmid-based genomic map of the 1.7 Mb *P.aerophilum* genome was constructed and used to identify 474 putative genes (16), but no homologues of the UNG or MUG/TDG UDG families were initially identified. Following the identification of *Tm*UDG, a novel UDG weakly related to *E.coli* MUG, in the thermophilic eubacterium *Thermotoga maritima,* a homologous ORF was identified in *P.aerophilum* encoding a new protein (designated *PaU*DG) with significant homology to *Tm*UDG (9). Here we show *PaU*DG to be an iron-sulphur protein with the characteristics of a 4Fe4S High Potential Iron Protein centre (HIPIP). Comparison of amino acid sequences and molecular modelling identified residues constituting the iron-sulphur cluster, and suggests this to be a common, though not universal, structural feature of the Family-4 UDGs.

EXPERIMENTAL PROCEDURES

Expression and purification of Pa-UDG

Pa-UDG was expressed in *E.coli* strain BL21 (DE3) pLysS from plasmid pET28-*Pa*-UDG essentially as described (9), with an N-terminal His_6 tag. The cell pellet was resuspended in buffer A (50 mM Tris pH 8, 100 mM NaCl, 10% glycerol), supplemented with 'Complete' EDTA free protease inhibitor cocktail (Roche), and stored at -20ºC. Cells were lysed by thawing, followed by a brief sonication on an ice / ethanol slurry (15 x 9s bursts with 9s cooling between bursts). The lysate was clarified by centrifugation at 50000xg and the supernatant was then incubated for 5 minutes at 80ºC to denature and precipitate the thermolabile *E.coli* proteins. The sample was cooled on ice, clarified by centrifugation at

50000xg then loaded onto a 5 ml Ni-NTA column pre-equilibrated in buffer A. The flow-through was discarded, as was a subsequent 10 column volume wash of buffer A supplemented with 10 mM imidazole. *Pa-U*DG was eluted in 5 column volumes of buffer A supplemented with 300 mM imidazole. The sample fractions were identified in the first instance by SDS-PAGE analysis (15% acrylamide), and subsequently by their yellow colour. Sample fractions were pooled, and their volume reduced (if required) to 10 ml by concentration in a Centriprep 20 spin concentrator (5 kD cut off) (Amicon). The sample buffer was then exchanged using a desalting column pre-equilibrated in buffer B (50 mM Sodium phosphate pH 7.5, 10 mM NaCl, 10% glycerol, 1 mM DTT, 'Complete' EDTA free protease inhibitors). A cation-exchange step was then used to complete the purification. During initial preparations, an HR5/5 Mono S column (Amersham-Pharmacia) was chosen, but during later preps an XK26/10 column packed with SP-Sepharose Fast Flow resin (Amersham-Pharmacia) was selected instead. Flow rates used were as recommended by the manufacturer for the column selected. In both cases the sample was applied to a column already equilibrated in buffer B. Both the flow-through and a 5 column volume buffer B wash were discarded. Bound protein was eluted via a linear NaCl gradient (10 – 500 mM) over 20 column volumes. The purified protein fractions were pooled and concentrated (as above), then transferred into buffer A supplemented with 1 mM DTT using a PD10 desalting column (BioRad). Purity was assessed by Coomassie stained SDS-PAGE (15% acrylamide), and the protein was stored in aliquots at -70ºC.

Spectroscopy

Ultra-violet/visible spectroscopy was carried out using a Shimadzu UV-2401PC recording spectrophotometer. Continuous Wave Electron Paramagnetic Resonance (CW-EPR) spectra

were obtained using a JEOL RE1X spectrometer equipped with an Oxford Instruments liquid helium cryostat. Samples were analysed as prepared, following reduction with sodium dithionite, and following oxidation with potassium ferricyanide.

RESULTS

The His_6-tagged *Pa*-UDG was overexpressed in BL21(DE3) cells using a pET28c(+)-*Pa-UDG* construct (9). The protein was purified from the cell lysate by means of heat treatment, immobilised metal-ion chromatography, and cation exchange chromatography to give an essentially pure sample migrating with an approximate molecular mass of 25 kD on SDS-PAGE (**Figure 1A**), while MALDI-TOF mass spectrometry gave a more precise mass of 24.248 kD (**Figure 1B**). Both results were consistent with the theoretical mass for His-tagged *Pa*-UDG (24.628 kD). N-terminal analysis of the purified protein prior to and following removal of the His_6-tag by digestion with thrombin confirmed its identity as *Pa*-UDG. Uracil-DNA glycosylase activity of the purified protein at 70°C was confirmed as described (6).

The pure protein was dialysed against a minimal buffer of 50 mM Tris-HCl pH 8.0, 100 mM NaCl and 1mM DTT for concentration and subsequent crystallographic analysis. The protein was highly soluble, and could be concentrated to > 30 mg ml^{-1}. Unexpectedly, dilute *Pa*-UDG (~1 mg ml^{-1}) was observed to be yellow in colour, and this colour intensified to dark olive and eventually brown as the sample was concentrated by ultrafiltration. The retention and concentration of the colour against a 5 kD cutoff membrane suggested a high molecular weight protein-associated chromophore rather than a small molecule contaminant. Consistent

with this, an adsorption spectrum of the concentrated protein displayed a broad peak around 370 - 400 nm, in addition to the normal absorption peaks around 280 nm, due to side chains of aromatic amino acid residues. (**Figure 2**). Absorption peaks in the 370 - 400 nm region can result from a variety of common biological chromophores, ranging from carotenes to porphyrins and iron-sulphur clusters. To determine whether any metals were present in the purified protein, the buffered sample was lyophilised and analysed by Inductively Coupled Plasma – Atomic Emission Spectroscopy (ICP – AES), which confirmed the presence of iron within the protein, with an estimated stoichiometry of ≈ 3 Fe atoms per mole of protein.

In order to ascertain the nature of the iron present in purified *Pa*-UDG, we recorded continuous wave electron-paramagnetic-resonance (CW-EPR) spectra of protein prepared using a Mono-S cation exchange step in the first instance (**FIGURE 3**). The EPR spectra of the enzyme clearly demonstrated the presence of iron sulphur centres in the sample. As prepared, the enzyme showed a weak spectrum characteristic of oxidised 3Fe4S centres at approximately g = 2.02. On reduction with sodium dithionite, this was replaced by a weak ferredoxin-like spectrum with peaks at approximately g = 2.06, 1.95 and 1.85. On oxidation with potassium ferricyanide a much stronger signal was observed, partly indicating an increase in the oxidised 3Fe4S centre, but also showing strong signals at g = 2.12 and 2.04 characteristic of an oxidised 4Fe4S High Potential Iron Protein (HPIP) centre, and demonstrating the presence of both 3Fe and 4Fe centres in the preparation.

Where a Mono-S column was used as the final purification step, the *Pa*-UDG eluted essentially as a single peak, although the chromatogram suggested that there may actually

be two peaks present, which had not been fully resolved. When the Mono-S column was replaced by an SP-Sepharose Fast Flow column *Pa*-UDG reproducibly eluted as two distinct peaks (**FIGURE 4A**), designated species 1 and 2. Each peak contained a pure coloured protein which migrated in SDS-PAGE with molecular weight consistent with *Pa*-UDG (**FIGURE 4B**) and was confirmed as such by MALDI-TOF mass spectrometry and Edman N-terminal sequencing in both cases (data not shown). The EPR spectra of the two peaks were very similar (**FIGURE 4C**). Both samples contained a 4Fe4S HIPIP centre, giving large signals in the oxidised state but little signal as prepared or in the reduced state. However, species 2 contained a small amount of the 3Fe centre observed in previously described preparations, whereas species 1 did not.

DISCUSSION

Taken together, the UV-visible, atomic absorption and electronparamagnetic resonance spectroscopic data all point to the presence of an integral cuboidal iron-sulphur cluster of the 4Fe4S HIPIP-type, in the Family-4 UDG from *Pyrobaculum aerophilum*. Depending on the methodology used in the purification of *Pa*-UDG, the 3Fe4S and 4Fe4S clusters may co-exist, and may be interconvertible as in other iron-sulphur proteins such as aconitase (17,18), although the 3Fe centres may be the result of damage or partial denaturation during the purification process. Resolution of two apparently compositionally identical 4Fe4S species in an ion-exchange column also suggests that multiple oxidation states with different net charges may also be possible. Cuboidal iron-sulphur clusters have previously been observed in DNA repair enzymes of the MutY/Nth/Ogg structural superfamily (19), such as the eubacterial Endonuclease III (Nth) (20) and the archaeal *Pa*-MIG, also identified in

Pyrobaculum aerophilum (21). However, to our knowledge, *Pa*-UDG is the first example of such a feature in the uracil-DNA glycosylase structural superfamily (14).

Location of cluster-ligand residues.

Iron-sulphur clusters of the HIPIP-type are usually attached via tetrahedrally directed bonds from the iron atoms to the Sγ atoms of four cysteine residues in the polypeptide chain. The *Pa*-UDG sequence contains six cysteine residues of which four are totally conserved in the characterised *Thermotoga maritima* and *Archeoglobus fulgidus* Family-4 UDGs (7,8), and in many homologous archaeal and eubacterial (putative) UDG sequences (**FIGURE 5**). These four cysteine residues are not totally conserved throughout Family-4 homologues, the first and third being replaced by aromatic residues in *Rickettsia* for example, nor are they restricted to hyperthermophiles, being present in Family-4 UDG homologues from spirochaetes, mycobacteria, *Clostridia* and *Deinococcus radiodurans*.

In previously-described HIPIP-type cuboidal iron-sulphur proteins, the sequence distribution of cysteine ligands varies considerably and consensus can only be obtained within protein families. The putative ligands in the Family-4 UDGs conform to a pattern : C-X_2-C-X_n-C-$X_{(14-17)}$-C, where 'n' ranges from 70-100. This is quite distinct from the Nth/MutY DNA repair enzymes, which show a much more localised consensus pattern : C-X_4PX-C-X_2-C-$X_{(6-8)}$-C, nor does it resemble any other known distributions of cysteine ligands in other iron-sulphur proteins characterised to date. If, as we suggest, these conserved cysteines act as ligands, then Pa-UDG must be able to fold so that the N-terminal C-X_2-C motif comes into sufficiently

close proximity to the central C-$X_{(14-17)}$-C motif, to bond to the iron atoms at the corners of the cuboidal 4Fe4S cluster.

To date, no structure for a Family-4 UDG has been reported. However, sequence threading and profile analysis techniques suggest that Family-4 UDGs will have a similar overall fold to the bacterial Family-2 MUG enzymes (14). Mapping the *Pa*-UDG sequence on to the crystal structure of *E.coli* MUG (12,13) locates the central pair of putative iron-sulphur cluster ligands on the surface exposed face of helix four and the loop that precedes it (**FIGURE 6A**). Cysteine residues at these positions (corresponding approximately to residues 72 and 87 in the MUG structure) would be well located to provide two ligands for a 4Fe4S cluster. The N-terminal C-X_2-C motif occurs in a segment of the *Pa*-UDG sequence that precedes the N-terminus of MUG, and topologically equivalent residues cannot therefore be located in the known MUG structure. However, the N-terminus of MUG is on the same face of the protein as the residues corresponding to the central cysteine pair in *Pa-U*DG. The N-terminal pair of putative 4Fe4S ligand residues occur 8 and 11 residues upstream of the residue in the *Pa-U*DG sequence that corresponds to the N-terminus of MUG, and would certainly be on the same face of the protein as the central pair of putative cluster ligands. Although the *E.coli* MUG protein lacks residues corresponding to this segment of *Pa-U*DG, the more distantly related Family-1 UDGs do possess corresponding segments of sequence. In Family-1 UDG structures this segment forms a turn and preceding helix that lies over the surface carrying the topological equivalents of MUG residues 71 and 87. If a similar structure were present in *Pa-U*DG, it would comfortably deliver the N-terminal C-X_2-C motif into a position suitable for providing the remaining pair of ligands for the iron-sulphur cluster. (**FIGURE 6B**)

Functional role of an iron-sulphur cluster

Iron-sulphur clusters occur in a wide range of enzymes, primarily as redox active co-factors participating directly in electron-transfer catalytic mechanisms. However, cuboidal 4Fe4S clusters have also been identified in non-redox enzymes, most notably in the Nth/MutY family of DNA repair enzymes (20,22,23). A variety of biochemical and biophysical studies suggest that the 4Fe4S cluster in these enzymes is not directly involved in catalysis (24). Instead, it functions as a structural 'cross-link' analogous to disulphide bonds or Zinc-fingers, which nonetheless contributes to substrate recognition by maintaining the structure of protein segments involved in DNA interactions (25-27). On the basis of the structural homology between the Famly-4 enzymes and the Family-2 bacterial MUG, the deduced site of the 4Fe4S cluster in *Pa-U*DG suggests that it would not participate directly in glycosylase activity. However, the central pair putative conserved cysteine ligands map to the beginning and end of a loop segment in MUG that is involved in contacts with the DNA phosphate backbone (**FIGURE 6C**) (12,13) so that, as in the Nth/MutY enzymes, the 4Fe4S cluster might probably play a role in substrate recognition but not catalysis. Determination of the precise role of the cuboidal 4Fe4S cluster in Family-4 uracil-DNA glycosylases must await the results of structural and mutagenesis studies, which are ongoing.

Acknowledgements

We are grateful to Renos Savva, Tracey Barrett and Bernard Connolly for useful discussions; to Angela Paul for assistance with sequencing and mass spectrometry, and Emile Brule for assistance with atomic absorption spectroscopy. This work was supported by the Cancer

Research Campaign (LHP). The generous financial support of the UBS to AAS and JJ is also gratefully acknowledged.

REFERENCES

1. Lindahl, T. (1993) *Nature* **362,** 709-715
2. Tye, B.-K., Chien, J., Lehman, I. R., Duncan, B. K., and Warner, H. R. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 233-237
3. Focher, F., Verri, A., Verzeletti, S., Mazzarello, P., and Spadari, S. (1992) *Chromosoma* **102(1,** S67-71.
4. Lindahl, T., and Nyberg, B. (1974) *Biochemistry* **13,** 3405-3410
5. Jacobs, K. L., and Grogan, D. W. (1997) *J Bacteriol* **179,** 3298-303.
6. Koulis, A., Cowan, D. A., Pearl, L. H., and Savva, R. (1996) *FEMS Microbiol. Lett.* **143,** 267-271
7. Sandigursky, M., and Franklin, W. A. (1999) *Curr. Biol.* **9,** 531-534
8. Sandigursky, M., and Franklin, W. A. (2000) *J. Biol. Chem.* **275,** 19146-19149
9. Sartori, A. A., Schar, P., Fitz-Gibbon, S., Miller, J. H., and Jiricny, J. (2001) *J Biol Chem* **276,** 29979-86.
10. Nedderman, P., and Jiricny, J. (1993) *J. Biol. Chem.* **268,** 21218-21224
11. Gallinari, P., and Jiricny, J. (1996) *Nature* **383,** 735-738
12. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J., and Pearl, L. H. (1998) *Cell* **92,** 117-129
13. Barrett, T. E., Schärer, O. D., Savva, R., Brown, T., Jiricny, J., Verdine, G. L., and Pearl, L. H. (1999) *EMBO J.* **18,** 6599-6609
14. Pearl, L. H. (2000) *Mut. Res. DNA Repair* **460,** 165-181
15. Volkl, P., Huber, R., Drobner, E., Rachel, R., Burggraf, S., Trincone, A., and Stetter, K. O. (1993) *Appl Environ Microbiol* **59,** 2918-26.

16. Fitz-Gibbon, S., Choi, A. J., Miller, J. H., Stetter, K. O., Simon, M. I., Swanson, R., and Kim, U. J. (1997) *Extremophiles* **1,** 36-51.

17. Kennedy, M. C., and Bienert, H. (1983) *J. Biol. Chem.* **263,** 8194-8198

18. Kennedy, M. C., Emptage, M. H., Dreyer, J.-L., and Bienert, H. (1983) *J. Biol. Chem.* **258,** 11098-11105

19. Nash, H. M., Bruner, S. D., Scharer, O. D., Kawate, T., Addona, T. A., Sponner, E., Lane, W. S., and Verdine, G. L. (1996) *Curr. Biol.* **6,** 968-980

20. Cunningham, R. P., Asahara, H., Bank, J. F., Scholes, C. P., Salerno, J. C., Surerus, K., Munck, E., McCracken, J., Peisach, J., and Emptage, M. H. (1989) *Biochemistry* **28,** 4450-4455

21. Yang, H., Fitz-Gibbon, S., Marcotte, E. M., Tai, J. H., Hyman, E. C., and Miller, J. H. (2000) *J Bacteriol* **182,** 1272-9.

22. Michaels, M. L., Pham, L., Ngheim, Y., Cruz, C., and Miller, J. H. (1990) *Nucleic Acids Res.* **18,** 3841-3845

23. Thayer, M. M., Ahern, H., Xing, D. X., Cunningham, R. P., and Tainer, J. A. (1995) *EMBO J.* **14,** 4108-4120

24. Fu, W. G., Ohandley, S., Cunningham, R. P., and Johnson, M. K. (1992) *J. Biol. Chem.* **267,** 16135-16137

25. Porello, S. L., Cannon, M. J., and David, S. S. (1998) *Biochemistry* **37,** 6465-6475

26. Golinelli, M. P., Chmiel, N. H., and David, S. S. (1999) *Biochemistry* **38,** 6997-7007

27. Kuo, C. F., McRee, D. E., Fisher, C. L., O'Handley, S. F., Cunningham, R. P., and Tainer, J. A. (1992) *Science* **258,** 434-440

FIGURE LEGENDS

Figure 1. Expression and purification of *Pa*-UDG

a) SDS-polyacrylamide gel showing fractions from various stages of purification. (1) Soluble fraction of *E.coli* cell lysate; (2) post heat-treatment; (3) unbound material from Ni-NTA resin; (4) 10 mM imidazol wash; (5) 300 mM imidazole eluate; (6) post-cation exchange chromatography. The protein after step 6 is > 95% pure.

b) MALDI-TOF mass spectrum of purified *Pa*-UDG (lane 6 above). The estimated peak mass of 24248 is consistent with the calculated mass of 24628 for the iron-free protein..

Figure 2. UV-visible Spectrum of *Pa*-UDG

In addition to the normal expected peak at 278 nm attributable to aromatic amino acid residues, the UV-visible spectrum of purified *Pa*-UDG shows an additional broad absorbance peak around 383 nm, giving the protein a yellow colour.

Figure 3. EPR-spectra of Mono-S purified *Pa*-UDG

EPR-spectra of *Pa*-UDG as prepared using a final Mono-S purification (**1**) step clearly shows a weak spectrum characteristic of oxidised 3Fe4S centres at approximately g = 2.02. On reduction with sodium dithionite (**2**) this spectrum was replaced by a weak ferredoxin-like spectrum with peaks at approximately g = 2.06 1.95 and 1.85. On oxidation with potassium ferricyanide (**3**) a much stronger signal was observed, partly indicating an increase in the oxidised 3Fe4S centre but, surprisingly, also showing strong signals at g = 2.12 and 2.04

characteristic of an oxidised 4Fe4S high potential iron protein centre. The spectra are most consistent with the simultaneous presence of both 3Fe and 4Fe centres in the preparation.

Figure 4. EPR-spectra of SP-Sepharose-purified *Pa*-UDG

a) Chromatogram of *Pa*-UDG eluting from an SP-Sepharose column with a shallow salt-gradient. Two species were clearly resolved.

b) SDS-PAGE of fractions from species 1 and species 2 shows identical molecular weight (confirmed by Mass spectrometry). A small amount (<5%) of a higher molecular weight contaminant is evident in species 2 but almost absent in species 1.

c) Purified species 1 as prepared showed no EPR signal (**1**), but developed a strong signal characteristic of an oxidised high potential iron protein (HIPIP) 4Fe4S centre on addition of the oxidant potassium ferricyanide (**1ox**). As prepared, the spectrum of purified Species 2 (**2**) still showed a weak signal above g=2.00 suggesting the presence of some oxidised 3Fe3S clusters, possibly reflecting damaged centres. Addition of ferricyanide to this protein also produced a strong characteristic HIPIP signal (**2ox**). Neither of these preparations showed any detectable g=1.94 signal due to the presence of the reducible 4Fe4S centre observed in the initial preparation. The EPR spectra of species 1 is fully consistent with only a single type of iron-sulphur centre.

Figure 5. Comparative Alignment of MUG/TDG and Family-4 UDG Sequences

Sequences of the MUG/TDG enzymes from human, mouse, fission yeast (schpo), *Escherichia coli*, *Serratia marcescens* (serma) and *Deinococcus radiodurans* (deira) in the top block are aligned with sequences of the Family-4 UDGs from *Pyrobaculum aerophilum* (pyrae), *Archeoglobus fulgidus* (arcfu), *Pyrococcus horikoshii* (pyrho), *Aquifex aeolicus*

(aquae), *Thermatoga maritima* (thema), *Halobacterium sp* (halob), *Trepanoma pallidum* (trepa), *Clostridium acetobutylicum* (cloac), *Deinococcus radiodurans* (deira), *Rickettsia prowazekii* (ricpr), *Streptomyces coelicolor* (strco) and *Mycobacterium tuberculosis* (myctu). Secondary structural elements observed in the crystal structure of *E.coli* MUG (12) or predicted in the Family-4 UDGs (14) are shown as light cylinders (α-helices) and dark arrows (β-strands). The N- and C-terminal motifs that form the active sites in the two families are boxed, and the cysteine residues proposed to act as ligands for the iron-sulphur cluster are shown in reverse. Segments of the amino acid sequences occurring significantly before the N-terminus or extending significantly beyond the C-terminus of *E.coli* MUG, are not shown.

Figure 6 Location of 4Fe4S cluster

a) Crystal structure of *E.coli* MUG. The N-terminus is indicated as a blue sphere, and the residues corresponding to the central pair of putative 4Fe4S cysteine ligands in the Family-4 UDGs are highlighted in magenta. Two orthogonal views are shown

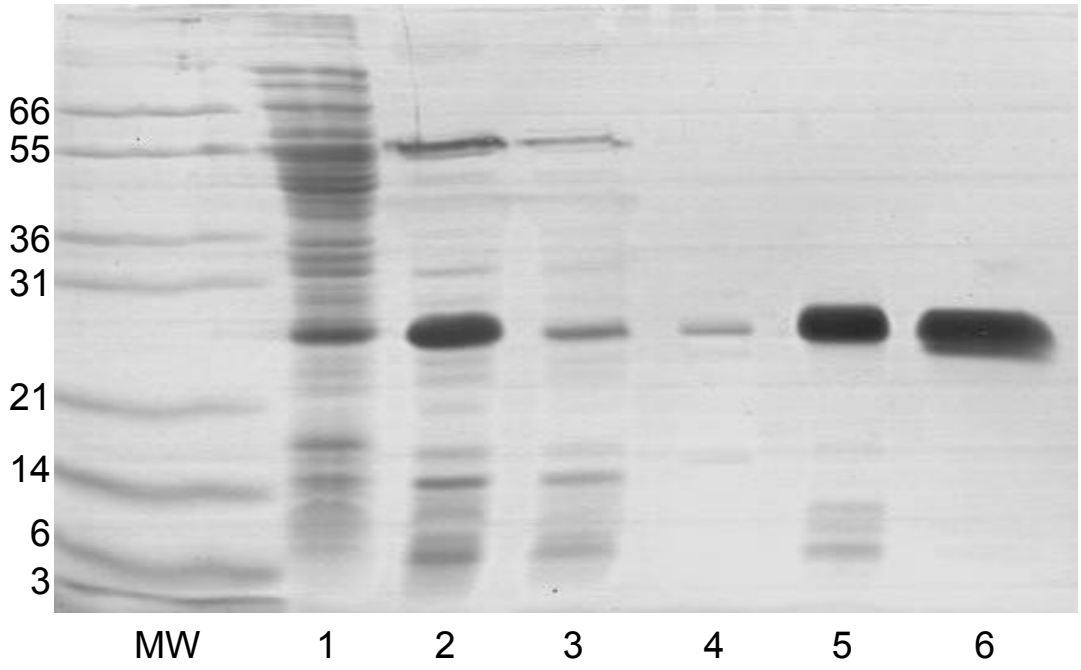
b) Hypothetical model of a 4Fe4S cluster bound to a Family-4 UDG. The cluster (grey CPK) is bound between the central pair of ligand cysteines that map into the MUG homology region (magenta) and the N-terminal pair of cysteines which are predicted to lie in an N-terminal helical extension on the MUG fold (silver).

c) Crystal structure of *E.coli* MUG bound to an oligonucleotide containing a non-hydrolysable uracil analogue. The polypeptide chain corresponding to the segment linking the central ligand pair in Family-4 UDGs is highlighted in green. If Family-4 UDGs interact with DNA in a similar manner to MUG, then this segment would be

involved in contacts with the DNA backbone, but would not play a direct role in catalysis of recognition of the scissile base.

Figure 1

a



b

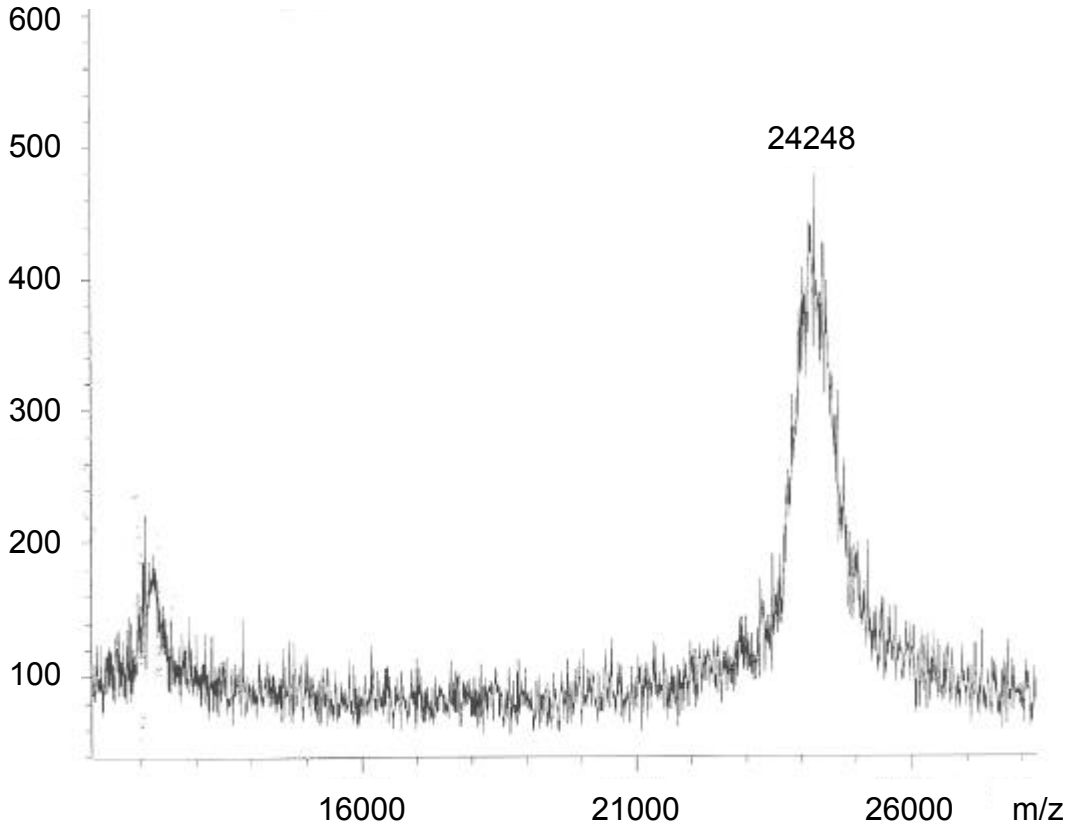


Figure 2

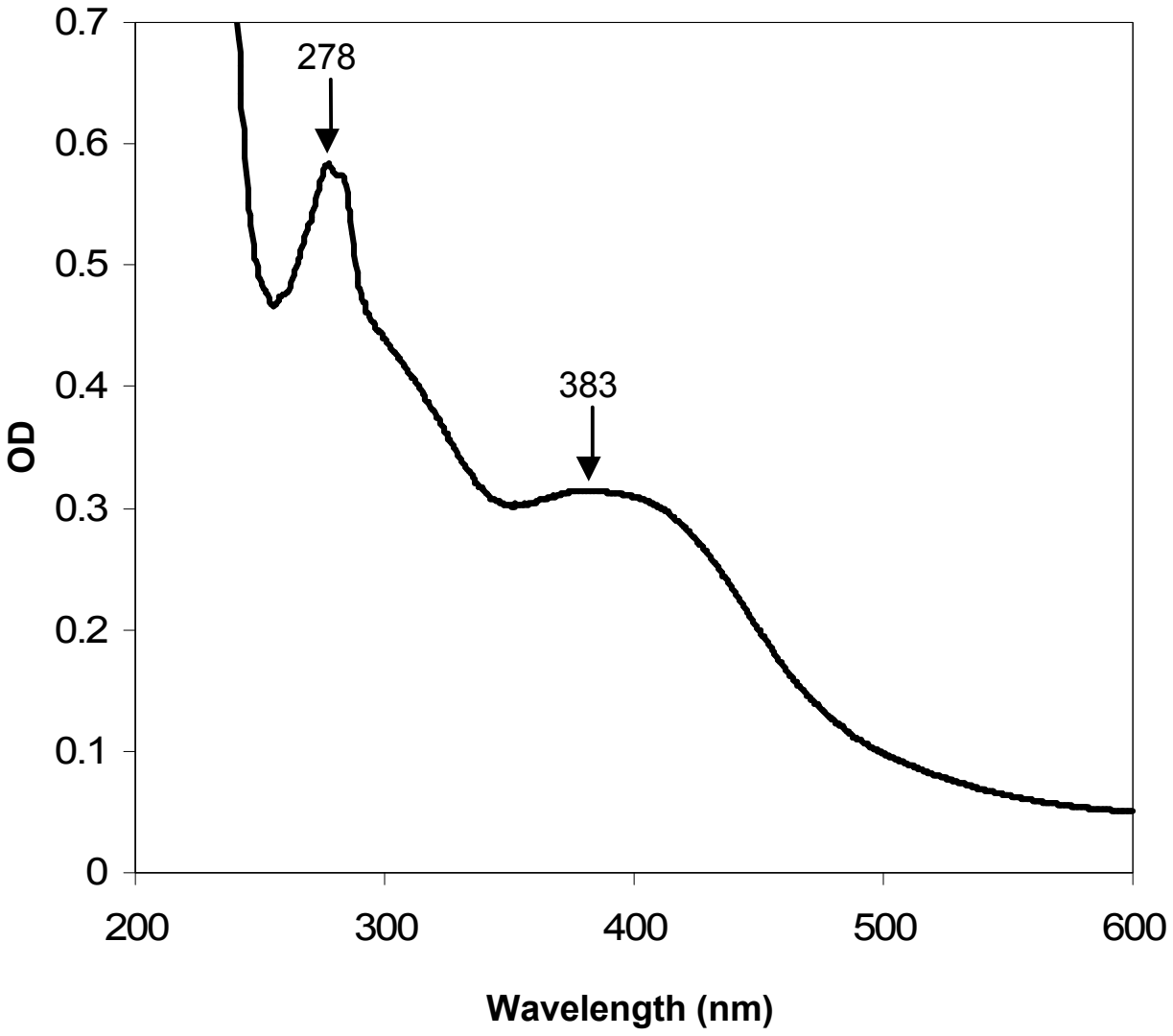


Figure 3

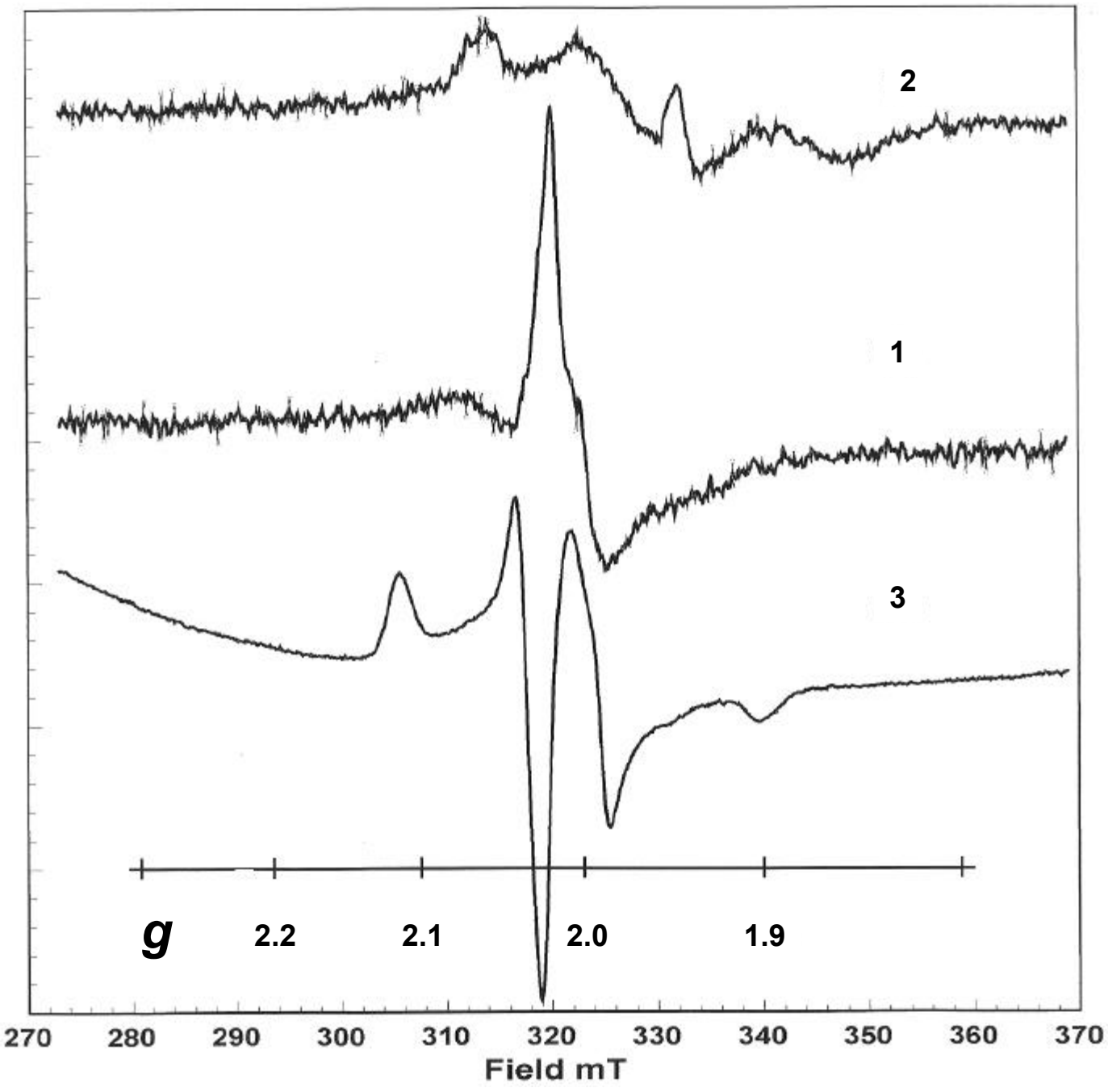
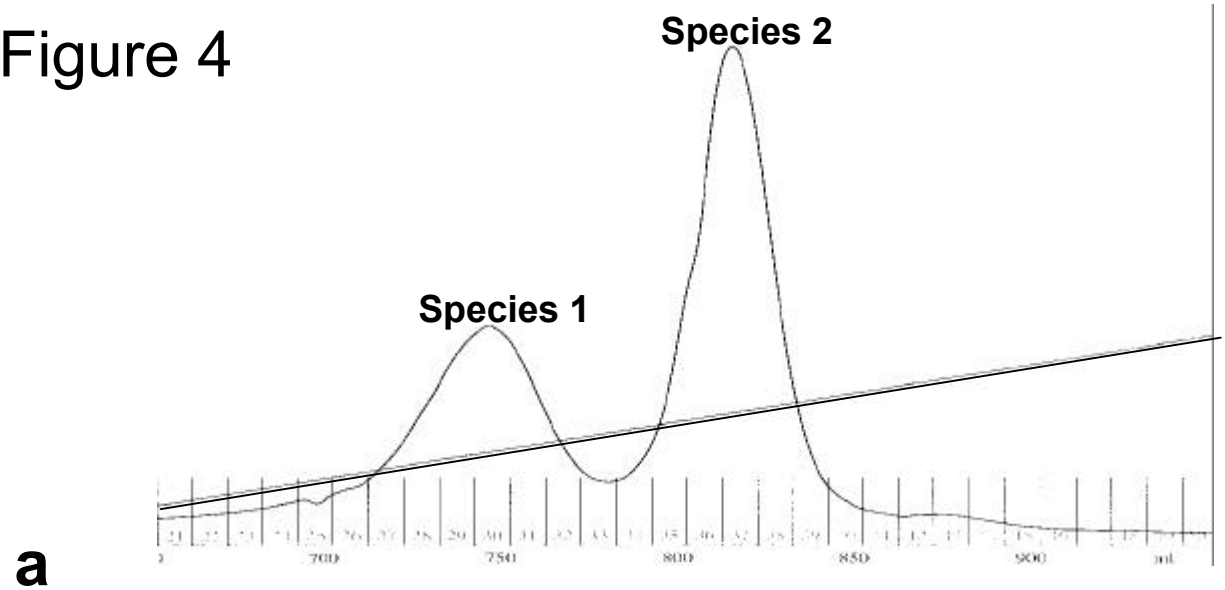
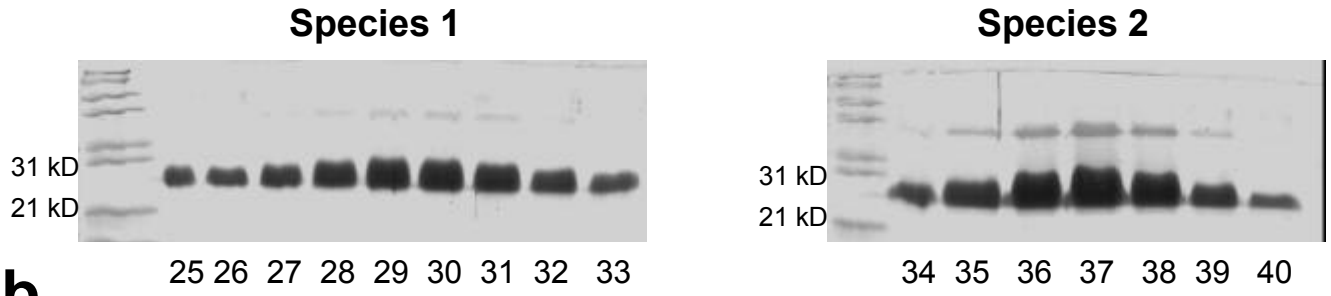


Figure 4

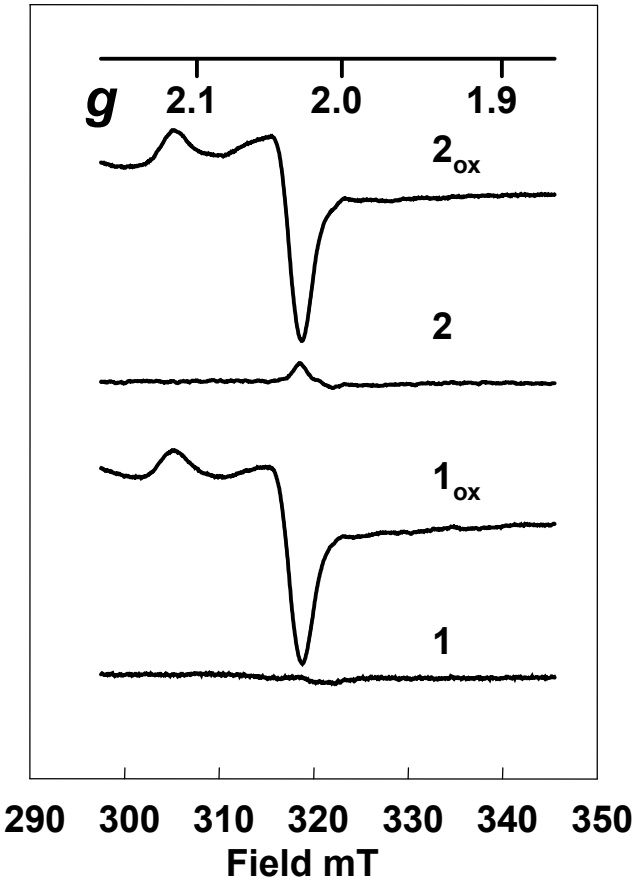


a



b

Fraction numbers



c

Figure 5

Figure 6